In [1]:

```
# Importing the libraries----

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [35]:

```
# Taking a closer look and counting the values associated with each Genre ----------

bestsellers_data['Genre'].value_counts()
```

Out[35]:

```
Non Fiction    310
Fiction        240
Name: Genre, dtype: int64
```

In [2]:

```
# Loading and reviewing the first five rows of the Dataframe------------------

bestsellers_data = pd.read_csv('bestsellers.csv')
bestsellers_data.head()
```

Out[2]:

	Name	Author	User Rating	Reviews	Price	Year	Genre
0	10-Day Green Smoothie Cleanse	JJ Smith	4.7	17350	8	2016	Non Fiction
1	11/22/63: A Novel	Stephen King	4.6	2052	22	2011	Fiction
2	12 Rules for Life: An Antidote to Chaos	Jordan B. Peterson	4.7	18979	15	2018	Non Fiction
3	1984 (Signet Classics)	George Orwell	4.7	21424	6	2017	Fiction
4	5,000 Awesome Facts (About Everything!) (Natio...	National Geographic Kids	4.8	7665	12	2019	Non Fiction

In [3]:

```
# Describing and taking summary statistics of the numerical values----------


bestsellers_data.describe()
```

Out[3]:

	User Rating	Reviews	Price	Year
count	550.000000	550.000000	550.000000	550.000000
mean	4.618364	11953.281818	13.100000	2014.000000
std	0.226980	11731.132017	10.842262	3.165156
min	3.300000	37.000000	0.000000	2009.000000
25%	4.500000	4058.000000	7.000000	2011.000000
50%	4.700000	8580.000000	11.000000	2014.000000
75%	4.800000	17253.250000	16.000000	2017.000000
max	4.900000	87841.000000	105.000000	2019.000000

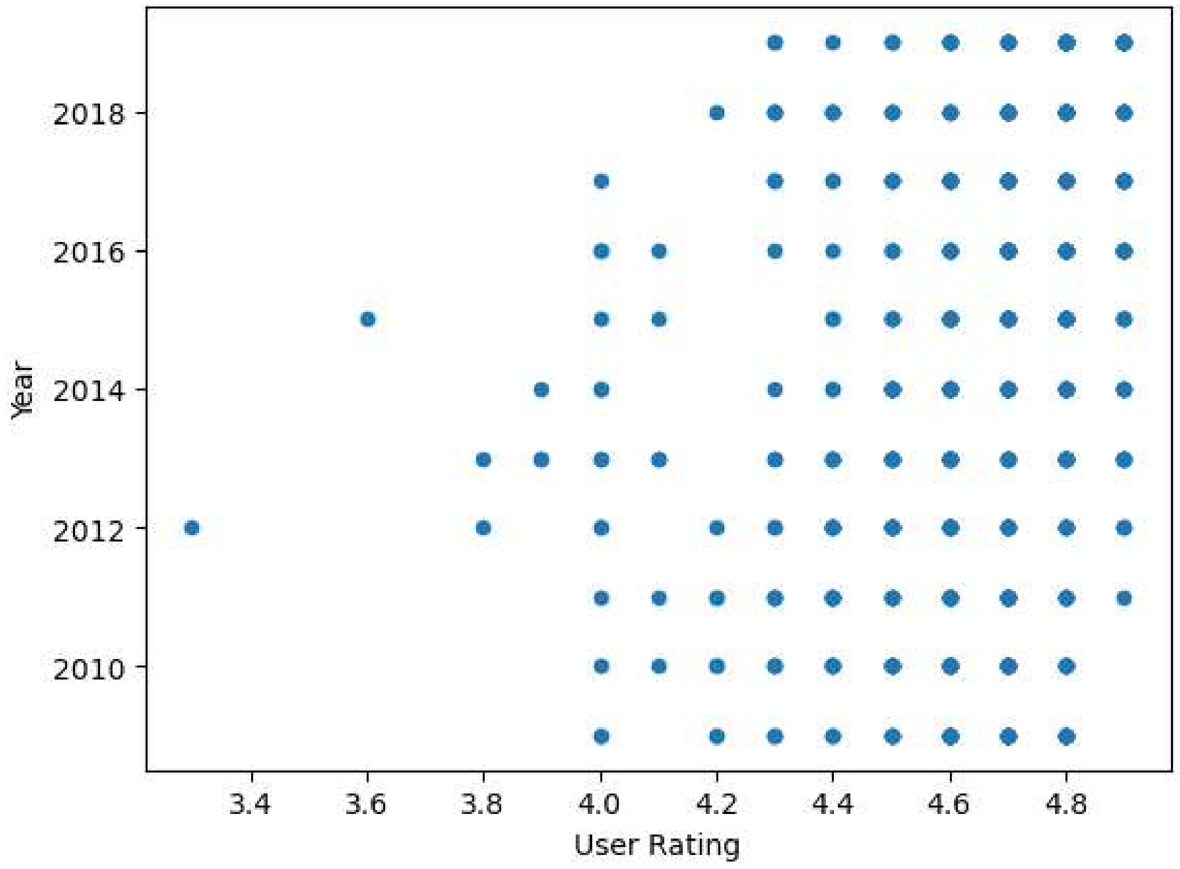
In [34]:

```
# Ploting a chart of the User Ratings by the various Years to know which year received the highest rating--------

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

bestsellers_data.plot(kind='scatter', x='User Rating', y='Year')

plt.show()
```



In [10]:

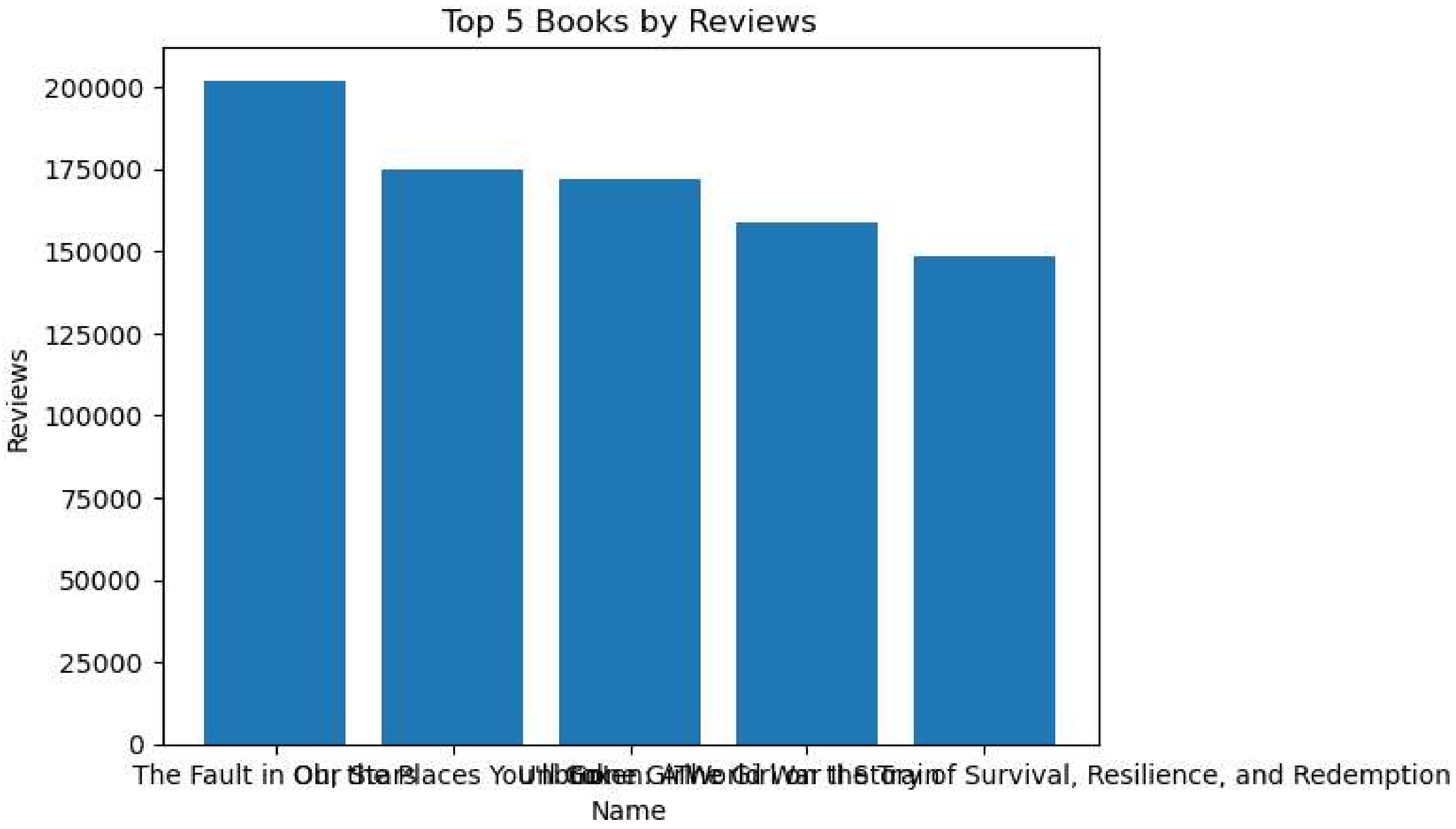
```
# Ploting a bar chart of the Top 5 Books by Reviews to know which book has the highest reviews--------


import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('bestsellers.csv')

reviews_by_book = df.groupby('Name')['Reviews'].sum().sort_values(ascending=False)
top_5_books = reviews_by_book[:5]

plt.bar(top_5_books.index, top_5_books.values)
plt.title('Top 5 Books by Reviews')
plt.xlabel('Name')
plt.ylabel('Reviews')
plt.show()
```



In [16]:

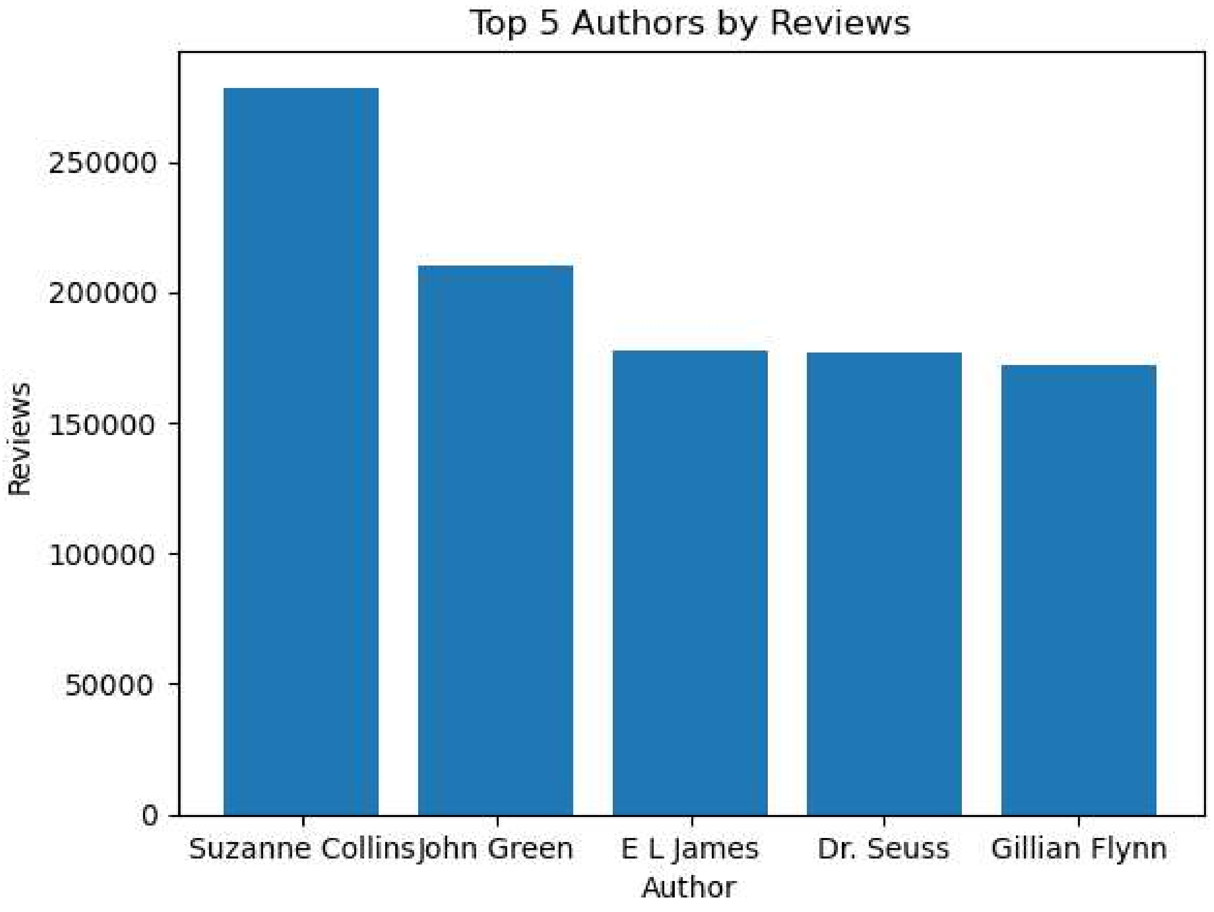
```
# Ploting a bar chart of the Top 5 Authors by Reviews to know which Author had the highest reviews--------

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('bestsellers.csv')
Authors_by_Reviews = df.groupby('Author')['Reviews'].sum().sort_values(ascending=False)

top_5_Authors = Authors_by_Reviews[:5]

plt.bar(top_5_Authors.index, top_5_Authors.values)
plt.title('Top 5 Authors by Reviews')
plt.xlabel('Author')
plt.ylabel('Reviews')
plt.show()
```



In [20]:

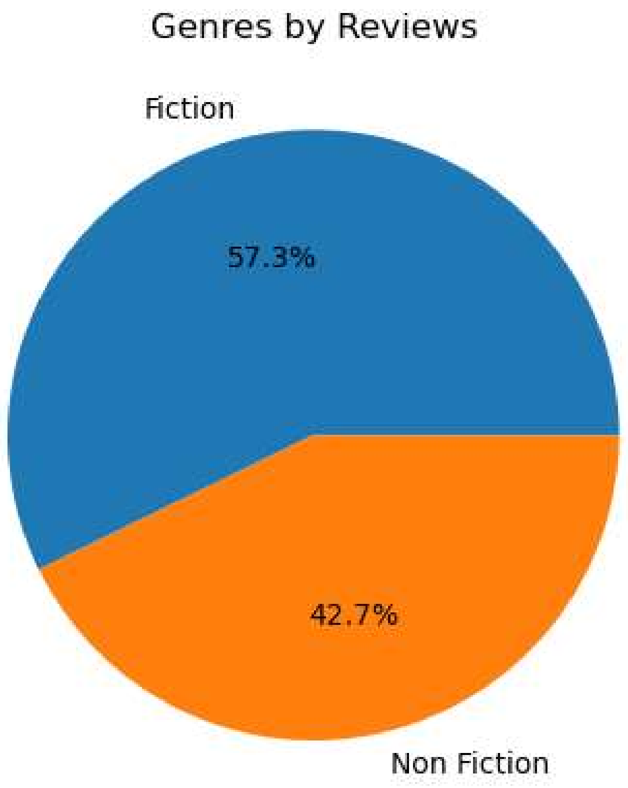
```
# Ploting a pie chart of the Genres by Reviews to know which of the genre received the highest reviews--------

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('bestsellers.csv')

genre_reviews = df.groupby('Genre')['Reviews'].sum().reset_index()

plt.pie(genre_reviews['Reviews'], labels=genre_reviews['Genre'], autopct='%1.1f%%')
plt.title('Genres by Reviews')
plt.show()
```



In [21]:

```
# Checking and printing duplicates if any

import pandas as pd
import seaborn as sns

df = pd.read_csv('bestsellers.csv')

duplicates = df[df.duplicated(['Name'])]

if duplicates.empty:
    print("No duplicates found")
else:
    print("Duplicates found:")
    print(duplicates)
```

```
Duplicates found:
                                                  Name            Author  \
10                           A Man Called Ove: A Novel   Fredrik Backman
21                         All the Light We Cannot See     Anthony Doerr
33                                            Becoming    Michelle Obama
36                            Between the World and Me  Ta-Nehisi Coates
41            Brown Bear, Brown Bear, What Do You See?   Bill Martin Jr.
..                                                 ...               ...
543                                             Wonder     R. J. Palacio
544                                             Wonder     R. J. Palacio
547  You Are a Badass: How to Stop Doubting Your Gr...       Jen Sincero
548  You Are a Badass: How to Stop Doubting Your Gr...       Jen Sincero
549  You Are a Badass: How to Stop Doubting Your Gr...       Jen Sincero

     User Rating  Reviews  Price  Year        Genre
10           4.6    23848      8  2017      Fiction
21           4.6    36348     14  2015      Fiction
33           4.8    61133     11  2019  Non Fiction
36           4.7    10070     13  2016  Non Fiction
41           4.9    14344      5  2019      Fiction
..           ...      ...    ...   ...          ...
543          4.8    21625      9  2016      Fiction
544          4.8    21625      9  2017      Fiction
547          4.7    14331      8  2017  Non Fiction
548          4.7    14331      8  2018  Non Fiction
549          4.7    14331      8  2019  Non Fiction

[199 rows x 7 columns]
```

In [22]:

```
# Checking if there are any missing data in the table------

import pandas as pd

data = pd.read_csv('bestsellers.csv')

missing_data = data.isnull().sum().sum()

if missing_data > 0:
    print("There are", missing_data, "missing values in the data.")
else:
    print("There are no missing values in the data.")
```

```
There are no missing values in the data.
```

In [36]:

```
# Checking the datatypes of each column-----

bestsellers_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 550 entries, 0 to 549
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Name         550 non-null    object 
 1   Author       550 non-null    object 
 2   User Rating  550 non-null    float64
 3   Reviews      550 non-null    int64  
 4   Price        550 non-null    int64  
 5   Year         550 non-null    int64  
 6   Genre        550 non-null    object 
dtypes: float64(1), int64(3), object(3)
memory usage: 30.2+ KB
```

In [ ]: